

## DOCUMENT RESUME

ED 293 875

TM 011 461

AUTHOR Vansickle, Timothy R.; Kapes, Jerome T.  
TITLE Equivalence of Computer-Based and Paper-Pencil Administrations of the Strong-Campbell Interest Inventory.  
PUB DATE Apr 88  
NOTE 15p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).  
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)  
EDRS PRICE MF01/PC01 Plus Postage.  
DESCRIPTORS College Students; Comparative Analysis; \*Computer Assisted Testing; Computer Uses in Education; Higher Education; Test Construction; \*Test Format; Testing; Test Reliability  
IDENTIFIERS \*Paper and Pencil Tests; \*Strong Campbell Interest Inventory

## ABSTRACT

First-, second-, and third-year students enrolled in an introductory educational psychology class at Texas A&M University in College Station were administered either a pencil-and-paper or computerized version of the Strong-Campbell Interest Inventory. The same or other version of the test was administered after 2 weeks. Focus is on equivalence of modality of test administration. Data were available for 75 students from four test conditions. With statistical significance as the standard, the results indicate that the two forms of the test are equivalent because reliabilities among the four student groups were similar, as were means and standard deviations. Very slight differences favored the computer-based test, with slightly higher reliability, a smaller standard deviation, and smaller difference scores, possibly resulting from the appearance of only one item on the monitor at a time. If these findings can be replicated, users of the computer-based instruments will benefit from increased reliability while maintaining confidence in established norms and validity. Three tables providing statistical data are appended. (SLD)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

ED 293875

EQUIVALENCE OF COMPUTER-BASED AND PAPER-PENCIL  
ADMINISTRATIONS OF THE STRONG-CAMPBELL INTEREST INVENTORY.

Timothy R. Vansickle  
Research Associate  
Educational Psychology  
Texas A&M University

and

Jerome T. Kapes  
Professor  
Industrial Vocational Technical Education/  
Educational Psychology  
Texas A&M University

Paper Presented at AERA Conference  
April 7, 1988 New Orleans, LA.

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

X This document has been reproduced as  
received from the person or organization  
originating it.

Minor changes have been made to improve  
reproduction quality.

Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OERI position or policy.

Running Head: SCII

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

TIMOTHY R. VANSICKLE

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

TM 011 461

### Introduction

The American Psychological Association recently published Guidelines for Computer-Based Tests and Interpretations (1986) which reiterates the importance of applying the same ethical, professional, and technical standards to computer-based tests as would be applied to traditional paper-pencil tests. Specifically, Guidelines 16 and 19 state that the equivalence of scores from computerized versions should be established and documented and comparison studies of computerized and conventional testing should be reported to establish the relative reliability of computerized assessment. This study was conducted to investigate the extent to which these guidelines were satisfied by the computer version of the Strong-Campbell Interest Inventory (SCII).

As early as 1974, studies were being conducted to investigate the equivalence of computer-based (CB) and paper-pencil (PP) administrations. Personality inventories were among the first assessment devices that were administered by computer using a remote terminal connected to a mainframe computer. Lushene, O'Neil, and Dunn (1974) completed an equivalency validity study of the MMPI and found that the booklet and computer-based administrations were not significantly different for female college students. An equivalency study on the California Psychological Inventory (Scissons, 1976) revealed that computer-based administration was a viable method for using this instrument with college students. No significant differences

emerged between the PP and CB administrations of the Eysenck Personality Inventory with mental health patients (Katz, 1981) and a sample of nurses (MacCullough & Mills, 1982). Finally, Harrell and Lombardo (1984) conducted an equivalency study with the 16PF and revealed no significant differences in score reliability and validity between the CB and PP versions.

While the majority of equivalency studies have been conducted on personality measures with college students, the use of computers in assessment has been found to produce results that are not significantly different from conventional test administrations with a variety of populations including; gifted and behavior problem children (Katz & Dalby, 1981), geriatric patients (Carr, Wilson, Ghosh, Ancill & Woods, 1982), physically handicapped (Wilson, Thompson, & Wylie, 1982), and mental health patients (Katz & Dalby, 1981). This variety of populations and instruments suggests that, at least for personality measures, CB versions are equivalent to PP administrations.

In the cognitive domain two recent studies have compared the CB and PP administrations of the Multidimensional Aptitude Battery (MAB) (Ackerson & Calder, no date; Bellantimo, 1985). Both studies revealed no significant differences between the two forms of the MAB.

In addition to use with the personality and cognitive measures cited above, the computer is now being used to administer interest inventories. This not only saves time, but appears cost effective and can provide immediate feedback to the

client which could make career counseling more efficient. O'Shea (1987) completed a study comparing the CB and PP administrations of the Harrington-O'Shea Career Decision Making System (CDM) with college freshmen. The results suggest that the two versions are equivalent.

The most widely used interest inventory in the United States is the Strong-Campbell Interest Inventory (SCII) (Zytowski & Warman, 1982). Its frequent use and availability of CB versions necessitates a study to determine the equivalence of the CB and the PP versions. The purpose of this study is to compare the two forms (CB & PP) of the SCII to satisfy APA Guidelines 16 and 19 concerning the equivalence of the computer version. Specifically, the question of equivalence will be addressed in differences between groups for the following: test-retest correlations, mean scores, standard deviations, and D-scores.

#### Method

##### Subjects

All subjects for this study were first, second, or third year students at Texas A&M University and were enrolled in an introductory Educational Psychology (Improvement of Learning 101) course. As part of the course requirements of EPSY 101, all students normally take the SCII with a discussion of the results following.

##### Instrumentation

Two versions of the SCII were administered; both the CB and PP versions were published by National Computer Systems (NCS).

This version was form T325 (1985 revision) which contains 325 items on a Like, Dislike, or Indifferent scale. The form used produces a profile consisting of three major parts: General Occupational Themes (6 scales), Basic Interest Scales (23 scales), and Occupational Scales (207 scales). This version of the SCII is the latest revision of the instrument and is sex balanced with matched occupational samples for 202 occupational scales and expanded "men in general" and "women in general" norm groups. All scale scores are T-scores with a mean of 50 and a standard deviation of 10. The NCS scanner and an IBM PC were used for scoring and production of the score profiles, as well as administering the computer-based SCII. A student demographic sheet was also administered which contained self-reported GPA, credit hours, major, and age.

#### Procedure

The study was conducted over a six week period of the spring semester 1987. During the study, score profiles and discussion of vocational interests were withheld until completion of the second administration of the SCII. Four intact sections (  $n = 100$  ) of Educational Psychology 101 students were selected from eight available sections, each of which had an enrollment of 25 students. The students were randomly assigned to one of the following four conditions within a two week test-retest design: group 1, PP version twice; group 2, PP followed by CB version; group 3, CB followed by PP version; and group 4, the CB version twice. The 75 subjects for whom complete data were available

were distributed as follows: group 1,  $n = 25$ ; group 2,  $n = 15$ ; group 3,  $n = 19$ ; and group 4,  $n = 16$ . Students failing to complete the second test were dropped from the study. Mortality in the CB groups occurred more frequently because of the individual nature of the CB modality, as these were administered individually by appointment while the PP versions were group administered.

Students received the standard instructions for the SCII with the additional instruction to perform on the second administration as they had on the first. After completing the two administrations of the SCII a class period was set aside for a group interpretation of the SCII and debriefing of the students.

### Results

The main question in this study was the relationship between the scores obtained on combinations of the two versions of the SCII. The mean and median test-retest coefficients are presented in Table 1. The mean test-retest coefficients for each group

-----  
Insert Table 1 about here  
-----

are as follows: group 1, 0.84; group 2, 0.84; group 3, 0.89; and group 4, 0.92.

A second question in the study concerns the differences between the four groups. A modified chi square statistic (Hakstian & Whalen, 1976) failed to provide evidence of

statistically significant differences between the four groups and yielded an  $M = 2.81$  with 3 df which was below that needed for significance ( $M = 7.81$  at the .05 level). However, if the  $N$  for each group remained at the starting level of 25 and the size of the differences remained the same statistically, significant results would have been obtained.

Further analysis dealt with differences among the means and standard deviations of the four groups. This analysis, presented in Table 2 used t-tests and analysis of variance procedures, and produced no significant difference between groups. However, the trend in the

-----  
 Insert Table 2 about here  
 -----

data indicated slightly smaller standard deviations on the second administration in all groups which would indicate that interests were solidifying and may be due to the greater familiarity with the test itself. Furthermore, the CB group had the smallest standard deviations for both the test and retest.

The last question to be answered concern differences between the "difference" or D-scores among the groups. All D-scores were computed using the absolute value of the scale scores on the second test minus the scale scores of the first test. The mean D-scores, by subscale, are presented in Table 3. From the table it can be seen that group 4 had the smallest D-scores (i.e., they had the smallest variation in scale scores between the two

administrations) but no trends are evident. None of the differences were significant at the .05 level.

-----  
Insert Table 3 about here  
-----

### Discussion

The driving force behind the research was the 1986 guidelines set forth by the APA. The guidelines cover issues of equivalence of CB tests and their PP counterparts. This study approached these issues using a systematic four group design which provided the needed control to determine equivalence of mode of administration.

Using statistical significance as the standard, the results would indicate that the two forms of administration for the SCII are equivalent because the reliabilities among the groups are similar as are the means, standard deviations, and mean D-scores. However, the M statistic would have provided a statistically significant difference for reliabilities in favor of the CB SCII if the groups would have remained balanced with and  $n = 25$  and the magnitude of differences in test-retest correlation coefficients remained approximately the same.

Although statistical significance is usually the criteria for accepting differences between groups, when addressing the question of equivalence of form, statistical significance may not be the appropriate standard. Many of the results indicate that there are slight differences which favor the use of CB

administrations, these include: slightly higher reliability, a smaller standard deviation, and smaller D-scores. This finding may be affected by the fact that in the CB SCII only one item is presented on the monitor at any one time. This could focus the examinee's attention on the task better than the booklet form of the PP administration. One of the observations made during the study (which was not part of the design and therefore exact time differences were not obtained) was time of administration. Those examinees taking the CB SCII took less time to complete the inventory and this could be a factor in increasing the reliabilities and decreasing the dispersion of the scores obtained in the CB mode.

This study is one of the first to address the issue of equivalence of modality of test administration for any instrument. The study provides some evidence for equivalence of the two modalities of the SCII administration but, also suggests that the CB SCII may be more reliable and faster. From this one small study there is reason to conduct further research with larger as well as with different populations, and different instruments. If these findings can be replicated, then users of CB instruments will benefit from the knowledge of increased reliability while maintaining confidence in previously established norms and validity.

## References

- Ackerman, K. & Calder, P. (no date). Comparison of the Wechsler Adult Intelligence-Revised and the Multidimensional Aptitude Battery.
- American Psychological Association. (1986). Guidelines for Computer-Based Tests and Interpretations. Washington, D.C.: Author.
- Bellantino, N. (1985). Computer versus paper-and-pencil administration of the Multidimensional Aptitude Battery, Ontario, Canada: University of Western Ontario.
- Carr, A. C., Wilson, S. L., Ghosh, A., Ancill, R. J., & Woods R. T. (1982). Automated testing of geriatric patients using a microcomputer-based system. Internationa Journal of Man-Machine Studies, 17, 297-300.
- Harrell, T. H., & Lombardo, T. A. (1984). Validation of an automated 16PF Administration Procedure. Journal of Personality Assessment, 48, 638-642.
- Hakstian, A. R., & Whalen, T. E. (1976). A k-sample significance test for independent alpha coefficients. Psychometrika, 41 (2), 219-231.
- Katz, L., & Dalby, J. T. (1981). Computer-assisted and traditional psychological assessment of elementary-school-aged children. Contemporary Educational Psychology, 6, 314-322.
- Katz, L. & Dalby, J. T. (1981). Computer and manual administration of the Eysenck Personality Inventory. Journal of Clinical Psychology, 37, 587-588.
- Lushene, R. E., O'Neil, H. F., & Dunn, T. (1974). Equivalent validity of a completely computerized MMPI. Journal of Personality Assessment, 38, 353-361.
- O'Shea, A. J. (1987). A comparison of microcomputer and paper-pencil career interest inventory administration. American Association for Counseling and Development.

- Ridgway, J., MacCullough, M. J., & Mills, H.E. (1982). Some experiences in administering a psychometric test with a light and pen and microcomputer. International Journal of Man-Machine Studies, 17, 265-278.
- Scissons, E. H. (1976). Computer administration of the California Psychological Inventory. Measurement and Evaluation in Guidance, 9, 24-30.
- Wilson, S. L., Thompson, J. A., & Wylie, G. (1982). Automated psychological testing for the severely physically handicapped. International Journal of Man-Machine Studies, 17, 291-296.
- Zytowski, D. G. & Warman, R. E. (1982). The changing use of tests in counseling. Measurement and Evaluation in Guidance, 15, 147-152.

Table 1. Median and mean test-retest correlations by group and instrument subscales.

|                              |        | Group 1 | Group 2  | Group 3  | Group 4  |
|------------------------------|--------|---------|----------|----------|----------|
| First Administration =       |        | Paper   | Paper    | Computer | Computer |
| Second Administration =      |        | Paper   | Computer | Paper    | Computer |
| -----                        |        |         |          |          |          |
| Subscales                    |        | n = 25  | n = 15   | n = 19   | n = 16   |
|                              |        |         |          |          |          |
| Administrative<br>(2 scales) | Median | .83     | .86      | .86      | .92      |
|                              | Mean   | .83     | .86      | .86      | .92      |
| GOT (6 scales)               | Median | .84     | .82      | .91      | .92      |
|                              | Mean   | .85     | .84      | .89      | .93      |
| BIS (23 scales)              | Median | .83     | .85      | .86      | .93      |
|                              | Mean   | .83     | .84      | .90      | .92      |
| OCC (119 scales)             | Median | .87     | .86      | .93      | .92      |
|                              | Mean   | .83     | .84      | .92      | .91      |
| All Scales<br>(150 scales)   | Median | .86     | .86      | .92      | .92      |
|                              | Mean   | .84     | .84      | .89      | .92      |

Table 2. Means and Standard Deviations by group and instrument subscales

| First Administration =<br>Second Administration =<br>-----<br>Subscales |      | Group 1<br>Paper<br>n = 25 | Group 2<br>Paper<br>Computer<br>n = 15 | Group 3<br>Computer<br>Paper<br>n = 19 | Group 4<br>Computer<br>Computer<br>n = 16 |
|-------------------------------------------------------------------------|------|----------------------------|----------------------------------------|----------------------------------------|-------------------------------------------|
| Administrative<br>(2 scales)                                            | Mean | 39.60                      | 38.70                                  | 39.27                                  | 38.88                                     |
|                                                                         | SD   | 12.84                      | 13.11                                  | 12.89                                  | 11.73                                     |
| GOT (6 scales)                                                          | Mean | 46.46                      | 52.40                                  | 49.50                                  | 51.82                                     |
|                                                                         | SD   | 10.17                      | 10.41                                  | 10.82                                  | 10.17                                     |
| BIS (23 scales)                                                         | Mean | 48.16                      | 52.39                                  | 50.32                                  | 52.51                                     |
|                                                                         | SD   | 10.01                      | 9.90                                   | 10.50                                  | 9.96                                      |
| OCC (119 scales)                                                        | Mean | 27.73                      | 28.90                                  | 28.23                                  | 29.00                                     |
|                                                                         | SD   | 12.08                      | 12.76                                  | 13.64                                  | 12.00                                     |
| All Scales<br>(150 scales)                                              | Mean | 40.59                      | 43.10                                  | 41.83                                  | 43.05                                     |
|                                                                         | SD   | 11.27                      | 11.54                                  | 11.73                                  | 11.10                                     |

Table 3. Mean D-scores by group and instrument subscales.

|                              | Group 1 | Group 2  | Group 3  | Group 4  |
|------------------------------|---------|----------|----------|----------|
| First Administration =       | Paper   | Paper    | Computer | Computer |
| Second Administration =      | Paper   | Computer | Paper    | Computer |
| -----                        |         |          |          |          |
| Subscales                    | n = 25  | n = 15   | n = 19   | n = 16   |
| Administrative<br>(2 scales) | 4.84    | 7.10     | 5.58     | 3.88     |
| GOT (6 scales)               | 3.70    | 6.04     | 3.78     | 3.41     |
| BIS (23 scales)              | 3.72    | 4.78     | 3.45     | 3.21     |
| OCC (119 scales)             | 4.30    | 5.29     | 4.19     | 4.17     |
| All Scales<br>(150 scales)   | 4.20    | 5.26     | 3.99     | 3.97     |